SYSTEM AND METHOD FOR OPTIMIZING INTERCONNECTIONS OF COMPONENTS IN A MULTICHIP MEMORY MODULE

TECHNICAL FIELD

The present invention relates generally to computer memory modules, and, more specifically, to methods and apparatus for improving signal integrity between a memory hub or other component on a memory module and memory devices contained on the memory module.

BACKGROUND OF THE INVENTION

A main focus of the contemporary semiconductor industry is the creation of smaller, faster, higher density, and more efficient memory modules. These efforts are often frustrated by cross talk and skew of signals being communicated on and to the memory modules, particularly as the memory modules become smaller. Cross talk is an inductive effect which can arise when a variable current flows through a conductor. Variable current creates a corresponding variable magnetic field surrounding the conductor capable of inducing a disruptive signal in any adjacent conductors passing through the magnetic field. As a consequence, the placement of conductors in a memory module must be carefully engineered in order to maintain suitable distances of separation between conductors to minimize the effects of cross talk.

Skew is a relatively fixed differential delay between two signals, commonly the result of the signals traveling different path lengths. One technique to eliminate skew is to make the path lengths along which signals are coupled the same length. In this way, signal travel time will be the same, thus eliminating any differential delay. Overall, the necessity of such careful considerations in both distancing conductors from each other and in creating equivalent path lengths to minimize the effects of cross talk and skew complicates efforts to produce effective memory modules with small dimensions.

Generally, memory modules are comprised of individual memory devices coupled in parallel on a circuit board. These memory devices can be dynamic random access memory ("DRAM") devices suitable for a wide variety of applications.

A partial top plan view of one type of memory module known in the art is shown in Figure 1. As illustrated, two registered dual in-line memory modules (DIMM) 100a, 100b include a plurality of memory devices 102-116 arranged on a circuit board 140 and connected by a command/address bus 142 to a register 144. The memory devices 102-116 and the conductors of the command/address bus 142 are situated on the circuit board 140 with enough space between them to minimize any cross talk. The register 144 receives command signals applied through a control bus 146 and address signals applied through an address bus 148 from an external memory controller (not shown), such as what is typically referred to as a "north bridge controller" in a conventional computer system.

As illustrated in the registered memory module 100 shown in Figure 1, the command signals applied to the register 144 include a row address strobe signal ("RAS#") (the "#" indicates the signal is active low), a column address strobe signal ("CAS#"), clock enable signals ("CKE0" and "CKE7"), a write enable signal ("WE#") and chip select signals ("CS0#"-"CS7#") to activate the DRAM devices 102-116 on the respective memory modules 100a, 100b. Other signals not latched by the register 144 include a clock ("CK0") signal, data signals ("DQ0-DQ63") corresponding to a 64-bit data word applied to the modules through a data bus 150, and a number of other signals that are not pertinent to the present discussion. In the registered DIMMs 100a, 100b, bank address signals ("B0-B7") corresponding to an 8-bit bank address and row/column address signals ("A0-A12") corresponding to a 13-bit address are also applied to the register 144 through the address bus 148. Typically, groups of the DRAM devices 102-116 are coupled to respective chip select signals CS0#-CS7#, with each group receiving a given chip select signal being designated a "rank" of memory. In the following discussion, each DIMM 100a, 100b is assumed to include memory devices 102-116 on both sides of the DIMM, and groups of 4 memory devices (*e.g.*, 102-108 and 110-116) are coupled to respective chip select signals CS0#-CS7# to define 4 ranks per DIMM. Each memory device 102-116 is thus a X16 device, meaning that each of the 4 memory devices in a given rank provides 16 of the 64 bit data bus DQ0-DQ63.

In operation, when a computer processor (not shown) reads data from, or writes data to, a specific memory address in a particular rank of memory device 102-

116, it sends a signal to the memory controller (not shown) over a host bus (also not shown). The request is analyzed by the memory controller, which applies corresponding address signals A0-A12 and the previously described command signals to the registered DIMMs 100a-b. These signals are latched into the registers 144 of both of the DIMMs 100a-b, with the latched chip select signals CS0#-CS7# determining the rank of memory that is accessed. Only one chip select signal CS0#CS7# is activated to access the corresponding rank of memory on one of the DIMMs 100a-b.

During write operations, the command signal includes address signals and command signals enabling the memory controller to access and write to a specific address in a respective rank of memory. Write data bits DQ0-DQ63 from the data bus 150 are then applied over an internal data path (not shown for the sake of clarity) on the DIMMs to the memory devices 102-116, and the memory devices in the active rank store the write data. The internal data path consists of individual traces running from the memory devices 102-116 to signal traces (not shown) on an edge of the circuit board 140. During write operations, the register 144 also operates to generate the appropriate command and timing signals to control the memory devices 102-116.

During read operations, the command signal includes address signals and command signals enabling the memory controller to access and read data from a specific address within the activated rank of memory. The read data stored in the active rank are then applied over the internal data path to the data bus 150 and, in turn, to the memory controller as read data bits DQ0-DQ64.

As can be seen in Figure 1, the off-module command and address signals are applied to the midpoint of the module 100 such that the length of the control bus 146 and the address bus 148 on the module 100 are short. However, since the memory devices 102-116 are disposed on either side of the register 144, the path lengths of the command/address bus 142 to the memory devices 102-116 are of different lengths. As a result, address and command signals coupled from the register 144 to the different memory devices 102-116 are susceptible to skew. For example, the difference in delay in coupling command and address signals from the register 144 to the memory devices 102 and 108 makes it difficult to capture the command and address signals at both

memory devices with a common clock signal. This potential for signal skew can seriously limit the operating speed of the memory devices 102-116.

One way to solve this problem is to increase the path lengths of the command/address bus 142 coupled to the memory devices 104-114 to make them equal to the path length of the command/address bus 142 to the devices 102 and 116. While such a solution is effective in alleviating skew, it requires the placement of a greater length of conductive lines on the DIMMs 100a, 100b. This consumes more space, increases propagation delay, and may adversely affect signal integrity. Further, as memory bus speeds continue to increase, a need will arise to buffer data signals along with the command address signals such that a data buffer will be included on each memory module 100a, 100b to perform a similar function for data signals as the register 144 does for command and address signals.

A new computer memory architecture currently being developed is known as a memory hub architecture. In a memory hub architecture, a system controller or memory controller is coupled over a high-speed data link, such as a fiber optic link, to several memory modules. The memory modules are typically coupled in a point-to-point or daisy chain architecture such that the memory modules are connected one to another in series. Each memory module includes a memory hub that is coupled to the corresponding high-speed data links and is also coupled to a number of memory devices on the module. The memory hubs efficiently route memory requests and responses between the controller and the memory devices via the high-speed data links. Computer systems employing this architecture can have a higher bandwidth because a processor can access one memory device or rank of memory while another memory device or rank is responding to a prior memory access. For example, the processor can output write data to one rank of memory in the system while another rank is preparing to provide read data to the processor.

The command, address, and data signals between each memory hub and the corresponding memory devices can experience cross talk and skew just as do the signals on conventional memory modules as previously discussed. To increase the overall bandwidth of a memory utilizing the memory hub architecture, the signals between the hub and memory devices are very high-speed, which only exacerbates the

problems created by any skew due to the more restrictive timing requirements, as will be understood by those skilled in the art. If each memory module has a layout like the DIMMs 100a, 100b of Figure 1, layout and routing congestion problems arise. With the memory hub being positioned in the center of the circuit board 140 in place of the register 144 and the DRAMs 102-116 positioned as shown, the skew of signals to and from each DRAM 102-116 is different. This skew presents timing problems for the memory hub, particularly with regard to read data from the DRAMs which will arrive at the hub at different times yet must be accurately captured. The hub could execute a synchronization process for each DRAM, but this would increase the complexity and cost of the memory hub. Routing congestion problems also arise with this layout due to all the signals that must be routed between the devices and the hub. While more layers could be added to the circuit board 1 40, this increases the complexity and cost of the board.

There is a need for a memory module that minimizes skew and maximizes signal integrity between a memory hub and memory devices as well as between the module and a memory controller.

SUMMARY OF THE INVENTION

The present invention is directed to a memory module and method for coupling memory devices contained on a memory module to a memory controller. According to one aspect of the present invention, a memory module includes a circuit board and a memory hub is positioned in approximately a center of the circuit board. A plurality of pairs of memory devices are positioned around the memory hub and arranged in pairs. Each memory device includes pins associated with a first functional group of signals adjacent a first end of the device and pins associated with a second functional group of signals adjacent a second end of the device. The first ends of the devices in each pair are positioned adjacent one another on the circuit board. An edge connector is positioned along an edge of the circuit board and coupled to the memory hub.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a block diagram of a portion of a conventional computer memory system containing a plurality of conventional DIMMs.

Figure 2 is a schematic top view of a memory module including a circuit board on which a number of memory devices are physically positioned around a memory hub according to one embodiment of the present invention.

Figure 3 is a top view of a conventional DRAM illustrating the physical location of data, address, and control pins.

Figure 4 is a schematic top view of a memory module similar to the memory module of Figure 2 but including dual edge connectors to reduce the pin count of each connector according to another embodiment of the present invention.

Figure 5 is a schematic top view of a memory module including a circuit board on which a number of wide data bus memory devices are physically positioned around a memory hub according to a further embodiment of the present invention.

Figure 6 is a schematic top view of a memory module including a circuit board on which a number of low-width DRAMs are physically positioned around a memory hub according to a further embodiment of the present invention.

Figure 7 is a block diagram of a computer system including a system memory having a memory hub architecture formed by one or more of the memory modules of Figures 2 and 4-6.

DETAILED DESCRIPTION OF THE INVENTION

Figure 2 is a schematic top view of a memory module 200 including a circuit board 202 on which eight memory devices 204 are physically positioned around a memory hub 206 and are physically oriented to minimize the skew of signals among the memory devices according to one embodiment of the present invention. The position and orientation of each memory device 204 is such that the electrical characteristics of conductive lines or busses interconnecting the memory device and the memory hub 206 are substantially the same for all memory devices, minimizing the skew of signals among the memory devices and thereby allowing for high-speed

operation of the memory module as will be described in more detail below. In the following description, certain details are set forth to provide a sufficient understanding of the present invention. One skilled in the art will understand, however, that the present invention may be practiced without these particular details.

The construction of the memory module 200 assumes a particular physical layout for pins or "pin out" for each of the memory devices 204. As a result, the assumed pin out for the memory devices 204 will first be described in more detail with reference to Figure 3, which is a top view of a conventional DRAM 300 illustrating the physical location of data, address, and control pins. Figure 3 shows that the data bus pins DQ0-7 are physically grouped toward one end 302 of the DRAM 300, while the control-address pins are physically grouped toward an opposite end 304. A pin 1 designator 306 in the upper left corner of the DRAM 300 is shown so that the orientation of the DRAM may be determined from the location of this pin 1 designator. This pin out is typical for high volume DRAMs and is assumed for the memory devices 204 of Figure 2 and for all other embodiments of the present invention described herein. The exact number and location of data and control-address pins may vary as long as data pins are grouped toward one end of the memory device 204 and control-address pins grouped toward another end of the DRAM.

Returning now to Figure 2, the memory hub 206 is coupled to each of the memory devices 204 through a respective data bus DQ. The DRAMs 204 are positioned in pairs, with each pair being located adjacent a given edge of the circuit board 202 and centered relative to the corresponding other two edges of the circuit board. For example, one pair of DRAMs 204 is positioned near the top edge of the circuit board 202 and centered relative to the left and right edges of the board. One DRAM 204 in each pair is rotated 180 degrees relative to the other, positioning the pin 1 designated edges of the DRAMs adjacent one another. By positioning the memory devices 204 in this configuration, the data bus DQ of each memory device has substantially identical electrical characteristics since each bus is approximately the same length. This minimizes skew among the data busses DQ as previously discussed.

The memory hub 206 is also coupled to each memory device 204 through a corresponding control-address bus CA. Once again, due to the physical positioning of the memory devices 204, the CA bus routed to each memory device has substantially the same electrical characteristics. This is true because, as shown in Figure 2, each CA bus is routed diagonally toward a corner of the board 202 and couples to the control-address pins of the memory devices 204, which are located opposite the pin 1 designator 306 and thus near the edges of the corresponding memory devices near the corners of the board 202.

The memory hub 206 is further coupled to an edge connector 207 positioned on a bottom edge of the circuit board 202 through control-address busses 208, 210 and data busses 212, 214. The busses 208-214 collectively form a "system bus" of the memory module 200 and couple the memory hub 206 to a high-speed data link (not shown). The dotted lines for the busses 208, 210 merely indicate that these busses may be routed under the corresponding memory devices 204. The layout of the memory hub 206 and memory devices 204 allows the busses 208-214 to be routed relatively directly from the edge connector 207 to the memory hub so that the lines forming each bus have relatively the same electrical characteristics, minimizing skew among signals within the busses, as will be appreciated by those skilled in the art.

In operation, the memory hub 206 receives memory requests from the high-speed data link (not shown) and, in response to such signals, applies address, data, and control signals to the memory devices 204 to thereby transfer data to and from the memory devices. The memory hub 206 initially processes downstream memory requests from a memory controller (not shown) directed to one of the memory modules 200 in a system memory to determine if the request is directed to the particular module, and only accesses the memory devices 204 when this is true. The hub 206 also processes upstream return requests from downstream memory modules 200, such as return requests including read data from a downstream module. As part of the processing of requests, the memory hub 206 translates requests from the high-speed data links into corresponding commands to properly access the memory devices 204, and also may include conversion circuitry to convert, for example, optical signals from

the high-speed data link into electrical signals. The architecture and operation of a system memory having a daisy-chain memory hub architecture including the memory module 200 will be described in more detail below with reference to Figure 7.

In one embodiment of the memory module 200, each of the memory devices 204 as a 9-bit data bus DQ and all of the memory devices are in the same rank. Accordingly, the DQ busses 212, 214 are each 36-bits wide to thereby form a 72-bit wide data bus of the memory module 200. In another embodiment, the memory module 200 includes two ranks of memory, with the second rank being formed by memory devices 204 (not shown) positioned on a back side of the circuit board 202 in the same way as just described for the memory devices on a front side of the board. More specifically, for each memory device 204 on the front side of the board 202 a corresponding device is positioned on the back side of the board with the same orientation (*i.e.*, the pin 1 designated ends of devices on the back are adjacent the pin 1 designated ends of corresponding devices on the front side).

By positioning the memory devices 204 and memory hub 206 in this way, the electrical characteristics of the data bus DQ routed to each memory device is substantially the same for all the data busses, reducing skew among data busses and thereby allowing higher speed operation of the memory module 200. The same is true for the control-address busses CA routed to each memory device 204. The positioning of the memory devices tool for and memory hub two of six also allows for relatively direct interconnection between the memory hub and to the edge connector 207 through the busses 208-214. As previously mentioned, this reduces skew among data lines in the busses 208-214 to thereby allow for higher speed data transfer between the hub and edge connector.

Figure 4 is a schematic top view of a memory module 400 according to another embodiment of the present invention. The memory module 400 is similar to the memory module 200 of Figure 2, but instead of the single edge connector 207 this memory module includes a lower edge connector 402 and an upper edge connector 404 positioned along a bottom and a top edge, respectively, of a circuit board 406. Components in the memory module 400 that are the same as previously described for

the memory module 200 of Figure 2 and given the same reference designations in Figure 4, and for the sake of brevity will not again be described in detail. The memory hub 206, memory devices 204, and the DQ and CA busses between the hub and memory devices are the same as corresponding components in the memory module 200.

In the memory module 400, a lower data bus 408 and lower control-address bus 410 are coupled between the lower edge connector 402 and the memory hub 406, while an upper data bus 412 and an upper control-address bus 414 are coupled between the upper edge connector 404 and the memory hub. Typically, half the lines of the overall bus or system bus of the memory module 400 would correspond to the lower data bus 408 and half to the upper data bus 412, and the same for the lower and upper control-address busses 410, 414. Recall, the memory module 400 would typically be connected to other modules in a daisy-chain architecture, and the use of the dual edge connectors 402, 404 enables one of the edge connectors to include the pins required for all signal lines of the system bus coupled to a downstream memory module and the other edge connector to include the pins required for all signal lines of the system bus coupled to an upstream memory module. The memory module 400 is useful in applications where the total number of pins required for the edge connector 207 may be more than can be economically or technically included in a single edge connector. The dual edge connectors 402, 404 may also simplify routing the signal lines for the system bus corresponding to the lines of the busses 408-414.

Figure 5 is a schematic top view of a memory module 500 including a circuit board 502 on which a number of wide data bus memory devices 504 are physically positioned around a memory hub 506 according to a further embodiment of the present invention. The memory module 500 is similar to the memory module 400 of Figure 4, but instead of including eight memory devices 204 positioned around the memory hub 206 the module 400 includes only four memory devices 504. In the memory module 500, each of the memory devices 504 is assumed to have a data bus DQ that is twice the width of the data bus DQ of the memory devices 204. As a result, only half the number of memory devices 504 is required, with the routing of the data busses DQ and control-address busses CA between the memory devices and the

memory hub 506 being similar to the corresponding memory devices on the memory modules 200 and 400. More specifically, the devices 504 are oriented such that the ends of each device near which the control-address pins are located (*i.e.,* the ends opposite the pin 1 designators) are adjacent a corresponding corner of the circuit board 502.

The memory module 500 includes a lower edge connector 508 and an upper edge connector 510 positioned along a bottom and a top edge, respectively, of the circuit board 502. The memory hub 506 is coupled through a lower data bus 512 and a lower control-address bus 516 to the edge connector 508, and through an upper data bus 514 and upper control-address bus 518 to the edge connector 510. Alternatively, in another embodiment the module 500 includes only the lower edge connector 508 and the both pairs of memory devices 504 are positioned on the bottom half of the circuit board 502. This allows the size of the circuit board 502 to be reduced to half the size if desired. Once again, if more than one rank is to be contained on the module 500, the additional memory devices 504 are positioned on a back side of the circuit board 502 in the same way. Cutting the number of memory devices 504 in the memory module 500 in half reduces the heat dissipation of the module. In one embodiment, each memory device 504 has an 18-bit wide data bus DQ so that the memory module 500 has a 72-bit wide system bus, which is the same as the memory modules 200 and 500.

Figure 6 is a schematic top view of a memory module 600 including a circuit board 602 on which a number of low-width memory devices 604 are physically positioned around a memory hub 606 according to a further embodiment of the present invention. The routing of data busses DQ and control-address busses CA between the hub 606 and devices 604 is analogous to that previously described for the memory module 400 of Figure 4, and the same is true of data busses 608, 610 and control-address busses 612, 614 that collectively for the module system bus coupled between the hub and edge connectors 616, 618. As a result, such interconnections and positioning of components will not again be described in detail. The module 600 further includes eight memory devices 604 on a back side of the circuit board 602, each

of these memory devices including an independent data bus DQ coupled to the memory hub 606, as illustrated by the dotted data busses DQ in Figure 6.

The memory module 600 would typically be utilized in applications where very high reliability of the module is required, such as in server systems where gigabytes of data may be stored in DRAM memory modules and the failure of one of the modules would result in the loss of significant amounts of data. With the module 600, each memory device 604 could be used to provide only a single bit of data to each data word being processed so that a failure of any of the individual memory devices may be corrected through error checking and correcting (ECC) techniques, as will be appreciated by those skilled in the art. Such ECC techniques can detect and correct single bit errors in a given data word but can only detect and not correct multiple bit errors. Thus, the reliability of the module 600 is improved, as required in many computer systems such as server systems as previously described. The memory hub 606 operates to properly address data stored in the memory devices 604 such that each device provides a single bit of a given data word on the system bus, as will be appreciated by those skilled in the art. In one embodiment, half the memory devices 604 have 4-bit data busses DQ and half have 5-bit data busses DQ as indicated in Figure 6.

Although the memory devices on each of the memory modules 200, 400, 500 and 600 are described as being DRAMs, other types of memory devices could also be utilized, as will be appreciated by those skilled in the art. Moreover, the number and orientation of memory devices on a given memory module will vary depending on the particular application for which the memory module is being designed, and the embodiments of Figures 2, 4, 5, and 6 are merely examples of a myriad of different physical layouts that are part of the present invention.

Figure 7 is a functional block diagram of a computer system 700 including a plurality of memory modules 702a-n corresponding to one or more of the memory modules 200 and 400-700 of Figures 2 and 4-7, respectively, according to one embodiment of the present invention. The memory modules 702a-n form a system memory 704 having a daisy-chain memory hub architecture as previously discussed.

The computer system 700 includes a processor 705 for performing various computing functions, such as executing specific software to perform specific calculations or tasks. The processor 705 includes a processor bus 706 that normally includes an address bus, a control bus, and a data bus. The processor bus 706 is typically coupled to cache memory 708, which, as previously mentioned, is usually static random access memory ("SRAM"). Finally, the processor bus 706 is coupled to a system controller 710, which is also sometimes referred to as a "North Bridge" or "memory controller."

The system controller 710 serves as a communications path to the processor 705 for a variety of other components. More specifically, the system controller 710 includes a graphics port that is typically coupled to a graphics controller 712, which is, in turn, coupled to a video terminal 714. The system controller 710 is also coupled to one or more input devices 718, such as a keyboard or a mouse, to allow an operator to interface with the computer system 100. Typically, the computer system 100 also includes one or more output devices 720, such as a printer, coupled to the processor 705 through the system controller 710. One or more data storage devices 724 are also typically coupled to the processor 705 through the system controller 710 to allow the processor 705 to store data or retrieve data from internal or external storage media (not shown). Examples of typical storage devices 724 include hard and floppy disks, tape cassettes, compact disk memories (CDs), and other types of fixed or removable storage media.

The system controller 710 is further coupled to the memory modules 702a-n in a point-to-point or daisy chain architecture through respective high-speed links 726 coupled between the modules and the system controller 710. More specifically, each memory module 702a-n includes a memory hub 728 coupled to corresponding high-speed links 726, where each memory hub 728 communicates over the corresponding high-speed links and controls access to a number of memory devices 730 contained on the memory module.

The high-speed links 726 may be optical, RF, or electrical communications paths, or may be some other suitable types of communications paths, as will be appreciated by those skilled in the art. In the event the high-speed links 734

are implemented as optical communications paths, each optical communication path may be in the form of one or more optical fibers, for example. In such a system, the system controller 710 and the memory modules 702a-n will each include an optical input/output port or separate input and output ports coupled to the corresponding optical communications paths.

Although the memory modules 702a-n are shown coupled to the system controller 710 in a daisy architecture, other topologies may also be used, such as a switching topology in which the system controller 710 is selectively coupled to each of the memory modules 702a-n through a switch (not shown), or a multi-drop architecture in which all of the memory modules 702a-n are coupled to a single high-speed link 726. Other topologies which may be used, such as a ring topology, will be apparent to those skilled in the art. One skilled in the art will also understand suitable circuitry for forming the memory hubs 206.

In the preceding description, certain details were set forth to provide a sufficient understanding of the present invention. One skilled in the art will appreciate, however, that the invention may be practiced without these particular details. Furthermore, one skilled in the art will appreciate that the example embodiments described above do not limit the scope of the present invention, and will also understand that various equivalent embodiments or combinations of the disclosed example embodiments are within the scope of the present invention. Illustrative examples set forth above are intended only to further illustrate certain details of the various embodiments, and should not be interpreted as limiting the scope of the present invention. Also, in the description above the operation of well known components has not been shown or described in detail to avoid unnecessarily obscuring the present invention. Finally, the invention is to be limited only by the appended claims, and is not limited to the described examples or embodiments of the invention.